ANALYSIS OF INTERACTIVE INTONATION IN UNACCOMPANIED SATB ENSEMBLES

Jiajie Dai, Simon Dixon
Centre for Digital Music, Queen Mary University of London, United Kingdom
{j.dai, s.e.dixon}@qmul.ac.uk

ABSTRACT

Unaccompanied ensemble singing is common in many musical cultures, yet it requires great skill for singers to listen to each other and adjust their pitch to stay in tune. The aim of this research is to investigate interaction in four-part (SATB) singing from the point of view of pitch accuracy (intonation). In particular we compare intonation accuracy of individual singers and collaborative ensembles. 20 participants (five groups of four) sang two pieces of music in three different listening conditions: solo, with one vocal part missing and with all vocal parts. After semi-automatic pitch extraction and manual correction, we annotated the recordings and calculated the pitch error, melodic interval error, harmonic interval error and note stability. We observed significant differences between individual and interactional intonation, more specifically: 1) Singing without the bass part has less mean absolute pitch error than singing with all vocal parts; 2) Mean absolute melodic interval error increases when participants can hear the other parts; 3) Mean absolute harmonic interval error is higher in the one-way interaction condition than the two-way interaction condition; and 4) Singers produce more stable notes when singing solo than with their partners.

1. INTRODUCTION AND BACKGROUND

Voice is our original instrument [8], even from prehistoric times [13], and it is one of the defining features of humanity [26]. This instrument communicates emotion, expressing joy and sadness, hope and despair. Throughout the history of vocal performance, various theories have been set forth on vocal aesthetics and intonation in both individual and ensemble settings. This paper investigates the influence of interaction between singers on the intonation of singing ensembles.

Intonation describes how a pitch is played or sung in tune [7]. Its extreme importance in Western music arises from the fact that it relates to both melody and harmony, two central aspects of tonal music. The accuracy of intonation is determined by culturally specific tuning systems such as the equal tempered tuning system in Western music [25].

Without interaction or accompaniment, it is extremely difficult to sing with accurate pitch. Only 0.01% of people have *absolute pitch* [22], which is the ability to identify or reproduce any given note on demand [2]. Others must rely on relative pitch for tuning, comparing current auditory feedback with the memory of recently heard tones. As this memory fades, singers may sing out of tune or exhibit pitch drift, where intonation moves away from the reference pitch during a performance [9, 12, 20]. Singers also use their muscle memory, a learnt relationship between muscle strength and pitch, to tune their pitch [1].

Although the intonation of singers in individual and group settings has been investigated, very little of this research addresses interaction between singers in vocal ensembles. In Western music, one common configuration for singing ensembles and choirs comprises four musical voices or parts: soprano, alto, tenor and bass (SATB); so we chose the SATB ensemble as the research target for this paper.

Music ensembles are well-characterised examples of interactive work groups [28]. Every member of a musical ensemble needs to execute his or her own part flawlessly as well as contribute to the overall performance in a manner that produces a cohesive, unified sound [3]. This means that individual singers have to stay in tune with their own part (their previous notes) and with other singers' parts (concurrent and previous notes) [18, p. 151]. This creates a practical difficulty for SATB singers, because they have multiple potentially conflicting reference pitches, as well as their own tonal reference, on which they could base their relative pitch, and attending to any specific one of these may be difficult.

Interaction plays an important role in ensemble performance, but its effects can be negative. Terasawa and Hiroko [23] claimed that the intonation accuracy of choral members was influenced by the progression of chord roots. Brandler and Peynircioglu [3] observed that participants learned new pieces of music more efficiently when learning it individually than with companions. Mürbe *et al.* [15] observed that singers' intonation accuracy is reduced in the absence of auditory feedback. When singers cannot hear themselves, they have to rely on their muscle memory to tune which leads to an inaccurate intonation. Dai and Dixon [4] noted that even the presence of an in-tune stimulus during singing reduced singers' accuracy.

Although many publications give guidelines to keep singers in tune by training them as excellent soloists [1,2], the interaction in SATB ensemble performance as it unfolds in real-time has not been fully researched. The target of this study is to test the influence of the various vocal parts and how the singers interact with each other, especially how hearing other singers influences the performance of each vocal part. These effects are tested in terms of their effect on intonation.

In the next section, we describe the research questions, hypotheses and experimental design. The methodology section follows, covering musical materials, experimental procedure and intonation metrics. Then in section 4 we present results in terms of pitch error, melodic interval error, harmonic interval error and note variability in different experimental conditions. This is followed by a discussion in section 5, and a conclusion in section 6. The recordings, annotated data and software are made freely available for research; details are given in section 8.

2. EXPERIMENTAL DESIGN

2.1 Research Questions and Hypotheses

This study of interactive intonation in unaccompanied SATB singing is driven by a number of research questions. Firstly, we wish to determine whether singers rely on a particular vocal part for intonation, which we test by systematically isolating each vocalist so that the other singers cannot hear them. We expect that the bass part, which often contains the root notes of chords, is more important as a tonal reference [23], leading to our first hypothesis: pitch error will be higher when the bass part is missing than when other voices are isolated.

The second research question involves the effect of hearing other voices on intonation. Previous work suggests that singers are distracted by simultaneous sounds when they are singing (see section 1), and they are less able to attend to their auditory feedback loop in order to sing accurately. This leads to hypothesis 2, that the conditions in which singers hear no other voice will have less melodic interval error than the conditions in which they hear other singers. This effect might be strengthened by conscious adjustment of singers to the other parts in order to improve the harmonic intervals. Thus as a corollary we frame our third hypothesis, that we expect to see less harmonic interval error when singers can hear each other than when they are isolated. An additional effect of interaction should be that singers adjust their pitch more during notes where they hear other singers (who might also be adjusting). Thus our fourth hypothesis is that within-note variability in pitch will be higher (note stability will be lower) when singers hear each other than when they do not.

2.2 Design

To test these hypotheses, a novel experiment was designed and implemented, by which we investigate the interaction between the four vocal parts. We define three different listening conditions, based on what the singer can hear as

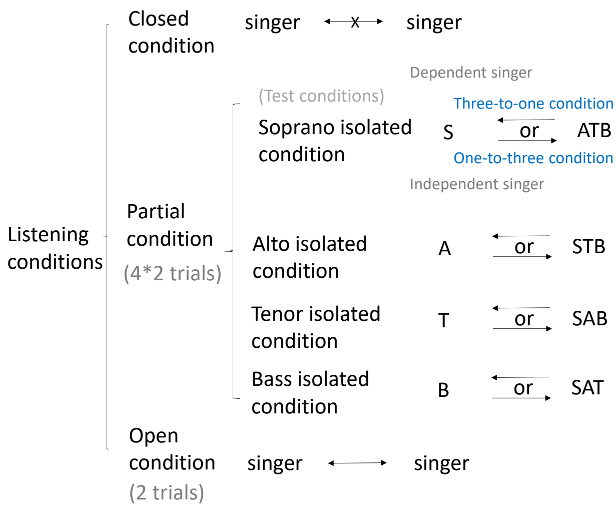

Figure 1: Listening and test conditions. The arrows indicate the direction of acoustic feedback.

they sing. In the *closed condition*, the singer hears no other voice than their own, thus they are effectively singing solo. In the *partially-open condition* (or *partial condition* for short), the singer can only hear some, but not all of the other vocal parts. This is achieved by isolating one singer from the other three, and allowing acoustic feedback (via microphones and loudspeakers) in one direction only, either from the isolated singer to the other three singers (*one-to-three condition*), or from the three singers to the isolated one (*three-to-one condition*). Finally, in the *open condition*, all singers can hear each other.

For testing the partial condition, there are four pairs of test conditions corresponding to the vocal part that is isolated and the direction of feedback. For example, one test condition is called the *soprano isolated one-to-three condition*, where the soprano sings in a closed condition, but all other parts hear each other (the soprano's voice being provided to the others via a loudspeaker). In such a case the isolated singer is called the *independent singer* as they are not able to react to the other vocal parts to choose their tuning. In other cases the singer can hear all (open condition) or some (partial condition) of the other voices, and thus is called a *dependent singer*. Figure 1 gives an overview of the listening and test conditions.

3. EXPERIMENTAL METHODS

3.1 Participants

20 adult amateur singers (10 male and 10 female) with choir experience volunteered to take part in the study. The age range was from 20 to 55 years old (mean: 27.95, median: 26.50, std.dev.: 7.84). Participants were compensated £10 for their participation. The participants were able to sing their parts comfortably and they were given the score and sample audio files at least 2 weeks before the experiment. They came from the music society and *a capella* society of the university and a local choir.

Training is a crucial factor for intonation accuracy. For testing the effect of training, all the participants were given a questionnaire based on the Goldsmiths Musical Sophistication Index [14]. The participants had an average of 3.3 years of music lessons and 5.8 years of singing experience.

3.2 Materials

Two contrasting musical pieces were selected for this study: a Bach chorale, "Oh Thou, of God the Father" (BWV 164/6) and Leo Mathisen's jazz song "To be or not to be". Both pieces were chosen for their wide range of harmonic intervals (see section 3.5.2): the first piece has 34 different harmonic intervals between parts and the second piece has 30 harmonic intervals. To control the duration of the experiment, we shortened the original score by deleting the repeat. We also reduced the tempo from that specified in the score, in order to make the pieces easier to sing and compensate for the limited time that the singers had to learn the pieces. The resulting duration of the first piece is 76 seconds and the second song is 100 seconds. Links to the score and training materials can be found in section 8.

The equipment included an SSL MADI-AX converter, five cardioid microphones and four loudspeakers. All the tracks were controlled and recorded by the software Logic Pro 10. The metronome and the four starting reference pitches were also given by Logic Pro. The total latency of the system is 4.9 ms (3.3 ms due to hardware and 1.6 ms from the software).

3.3 Procedure

A pilot experiment with singers not involved in the study was performed to test the experimental setup and minimise potential problems such as bleed between microphones. Then the participants in the study were distributed into 5 groups according to their voice type, time availability and collaborative experience (the singers from the same music society were placed in the same group). Each group contained two female singers (soprano and alto) and two male singers (tenor and bass). Each participant had at least two hours practice before the recording, sometimes on separate days. They were informed about the goal of the study, to investigate interactive intonation in SATB singing, and they were asked to sing their best in all circumstances.

For each trial, the singers were played their starting notes before commencing the trial, and a metronome accompanied the singing to ensure that the same tempo was used by all groups. Each piece was sung 10 times by each group. The first and the last trial were recorded in the open condition. The partial and closed condition trials, consisting of 8 test conditions, 4 (isolated voice) $\times$ 2 (direction of feedback), were recorded in between. The order of isolated conditions was randomly chosen to control for any learning effect. For each isolated condition, the three-to-one condition always preceded the one-to-three condition. We use the performance of isolated singers in the one-to-three conditions as the data for the closed condition.

The singers were recorded in two acoustically isolated rooms. For the partial and closed conditions, the isolated singers were recorded in a separate room from the other three singers. Loudspeakers in each room provided acoustic feedback according to the test condition. There was no visual contact between singers in different rooms. With the exception of warm-up and rehearsal, but including all the trials and the questionnaire, the total duration of the experiment for each group was about one hour and a half.

3.4 Annotation

The experimental data comprises 5 (groups) $\times$ 4 (singers) $\times$ 2 (pieces) $\times$ 10 (trials) = 400 audio files, each containing 65 to 116 notes. The software *Tony* [10] was chosen as the annotation tool. *Tony* performs pitch detection using the PYIN algorithm, which outperforms the YIN algorithm [11], and then automatically segments pitch trajectories into note objects, and provides a convenient interface for manual checking and correction of the resulting annotations. For each audio file, we exported two `.csv` files, one containing the note-level information (for calculating pitch and interval errors) and the other containing the pitch trajectories (for calculating pitch variability). All the intonations were measured by twelve-tone equal temperament, expressed in semitones according to MIDI standard pitch numbering. It took about 67 hours to manually check and correct the 400 files, resulting in 49200 annotated single notes, to which we added information on the singer (anonymised), score notes and metrics of accuracy.

3.5 Intonation Metrics

To quantify the effects of interaction on intonation, we measure pitch accuracy in terms of pitch error, melodic interval error, harmonic interval error and note stability, defined below.

3.5.1 Pitch Error

Assuming that a reference pitch has been given, *pitch error* can be defined as the difference between observed pitch and score pitch [12]:

$$e_i^p = \bar{p}_i - p_i^s \tag{1}$$

where $\bar{p}_i$ is the median of the observed pitch trajectory of note i (calculated over the duration of an individual note), and p_i^s is the score pitch of note i.

To evaluate the pitch accuracy of a sung part, we use *mean absolute pitch error* (MAPE) as the measurement. For a group of M notes with pitch errors $e_1^p, \ldots, e_M^p$, the MAPE is defined as:

$$\text{MAPE} = \frac{1}{M}\sum_{i=1}^{M} |e_i^p| \tag{2}$$

3.5.2 Melodic and Harmonic Interval Error

A musical *interval* is the difference between two pitches [19], which is proportional to the logarithm of the ratio of the fundamental frequencies of the two pitches. We distinguish two types of interval in this experiment: in a *melodic interval*, the two notes are sounded in succession; while in a *harmonic interval*, both notes are played simultaneously (Figure 2).

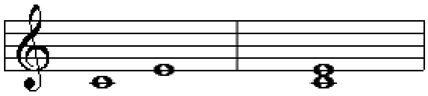

Figure 2: A melodic interval and harmonic interval of a major third (four semitones).

We thus calculate the melodic interval error as the difference between the observed and score intervals:

$$e_i^m = (\bar{p}_{i+1} - \bar{p}_i) - (p_{i+1}^s - p_i^s) \tag{3}$$

where p_i^s and p_{i+1}^s are the score pitches of two sequenced notes, and $\bar{p}_i$ and $\bar{p}_{i+1}$ are their observed median pitches. Similarly, harmonic interval error is defined as:

$$e_{i,A,j,B}^h = (\bar{p_{i,A}} - \bar{p}_{j,B}) - (p_{i,A}^s - p_{j,B}^s) \tag{4}$$

where $p_{i,A}^s$ and $p_{j,B}^s$ are the score pitches of two simultaneous notes from singers A and B respectively, and $\bar{p}_{i,A}$ and $\bar{p}_{j,B}$ are their observed median pitches.

The *mean absolute melodic interval error* (MAMIE) for M intervals is calculated as follows:

$$\text{MAMIE} = \frac{1}{M}\sum_{i=1}^{M} |e_i^m|. \tag{5}$$

The *mean absolute harmonic interval error* (MAHIE) is calculated similarly (where we simplify the notation and assume M harmonic intervals in total, indexed by i):

$$\text{MAHIE} = \frac{1}{M}\sum_{i=1}^{M} |e_i^h|. \tag{6}$$

Harmonic intervals were evaluated for all pairs of notes which overlap in time. If one singer sings two notes while the second singer holds one note in the same time period, two harmonic intervals are observed. Thus indices i and j in Eq. (4) are not assumed to be equal.

3.5.3 Note Stability

Pitch stability has been defined as the mean square pitch error of the note trajectory [17, 24], annotated using a fine time resolution, in this case Tony's default hop size of 5.8ms (section 3.4). We prefer to call this *pitch variability*, as higher values correspond to less stable notes. For a note trajectory for note i consisting of N frames, if the pitch of frame n is $p_{i,n}^f$ and the median pitch $\bar{p}_i$, the note variability ν_i is given by:

$$\nu_i = \frac{1}{N}\sum_{n=1}^{N} |p_{i,n}^f - \bar{p}_i|^2 \tag{7}$$

The *mean note variability* (MNV) is the mean variability of M notes:

$$\text{MNV} = \frac{1}{M}\sum_{i=1}^{M} \nu_i \tag{8}$$

4. RESULTS

The primary aim of this study was to test experimentally whether, and under what conditions, interaction is beneficial or detrimental to SATB intonation accuracy. We tested the intonation accuracy of individuals by pitch error (section 4.1), melodic interval error (section 4.2) and note stability (section 4.4); and tested the intonation of pairs of singers by harmonic interval error (section 4.3). In order to avoid biasing mean errors by outliers, where a participant sang a wrong note rather than an out-of-tune attempt at the correct pitch, all the tests exclude notes with pitch error or interval error larger in magnitude than one semitone. 96.4% of observed notes had an absolute pitch error less than one semitone.

4.1 Pitch Error

The first task is to investigate whether the ensemble depends on a certain vocal part to tune their pitch. After excluding the notes which have an absolute pitch error larger than one semitone (3.6%), most of the observed notes are relatively accurate (mean: 0.25 semitones; median: 0.26; std.dev.: 0.07).

We compute pitch error for the three non-isolated singers in each three-to-one condition and open condition, and analyse results by test condition. The MAPE was computed as an average across the three non-isolated singers and the five groups. For example, in the soprano isolated three-to-one condition, we average the pitch errors of alto, tenor, bass parts from each group and report the resulting MAPE. We compare these results with the performance of the same three singers in the open conditions.

A correlated samples analysis of variance (ANOVA) showed a significant difference in MAPE between three-to-one and open conditions ($F(1,21625)=13$, $p<.001$). The MAPE of the three-to-one condition is less than the MAPE of the open condition. We then performed separate ANOVAs for each isolated voice type (Table 1), and found that the results vary across test conditions. The bass and tenor isolated three-to-one conditions both showed significant differences, while the results for the other two voice types were not significant.

Test condition	Partial vs open condition
Soprano isolated	$F(1,9391)=2.86$, $p=0.09$
Alto isolated	$F(1,9614)=0.61$, $p=0.11$
Tenor isolated	$F(1,9742)=5.07$, $p=0.02$*
Bass isolated	$F(1,10223)=14.39$, $p<.001$***

Table 1: Results of correlated samples ANOVAs for three-to-one and open listening conditions (***p<.001; **p<.01; *p<.05))

These results suggest that the bass part is the most influential vocal part in all observed groups. However, the direction of influence is the opposite of that hypothesised: removing the bass vocal part from the ensemble reduces the observed pitch error on average.

The next ANOVA shows that the MAPE is significantly different between the test conditions in the three-to-one listening condition (F(3,12948)=28.67, p<.001). Table 2 shows the 95% confidence intervals, which demonstrate that the bass and tenor isolated conditions are significantly different from all other three-to-one conditions. The bass isolated condition has 4 cents MAPE less than soprano and alto isolated conditions, and 2 cents MAPE smaller than the tenor isolated condition.

Test condition	MAPE	Confidence interval
Soprano isolated	0.2484	[0.2420, 0.2548]
Alto isolated	0.2483	[0.2422, 0.2545]
Tenor isolated	0.2328	[0.2271, 0.2385]
Bass isolated	0.2082	[0.2028, 0.2135]

Table 2: Mean absolute pitch error (MAPE) and 95% confidence intervals for three-to-one test conditions, for all non-isolated singers and all groups.

These results contradict hypothesis one: when singers do not hear the bass part, they sing more accurately on average, as shown by comparisons within the three-to-one conditions and between the three-to-one and open conditions.

4.2 Melodic Interval Error

To test the influence of interaction on adjacent notes within a voice (hypothesis two), melodic interval error was calculated. 91.9% of the note pairs have a melodic interval error smaller than one semitone (mean:0.21; median:0.21; std.dev.:0.07).

We performed a correlated-samples ANOVA to test the effect of listening condition on MAMIE. The MAMIE is significantly different across listening conditions (F(2,18333)=27.96, p<.001). The listening condition of singing without hearing any partners (closed) has smaller MAMIE than the listening conditions with partners (partial and open). Table 3 shows the mean and confidence intervals for the three listening conditions where the closed listening condition has 3 cents smaller MAMIE than the open listening condition.

Listening condition	MAMIE	Confidence interval
Closed condition	0.1874	[0.1828, 0.1919]
Partial condition	0.2001	[0.1953, 0.2049]
Open condition	0.2138	[0.2102, 0.2174]

Table 3: Mean absolute melodic interval error (MAMIE) and 95% confidence intervals for each listening condition.

The acoustic feedback from other vocal parts increases MAMIE, which concurs with findings from previous research [15] and supports hypothesis two. The accompaniment from other vocal parts may mask the singer's own voice or distract the singer's attention from their own intonation. Alternatively, the increase in melodic interval error could be a side effect of deliberate adjustment of intonation to reduce harmonic interval error.

4.3 Harmonic Interval Error

Beside the intonation accuracy of individual singers, the accuracy of pairs of singers was also tested. There are four individual singers and up to six harmonic intervals simultaneously present at any point in time. All the harmonic intervals were observed under two circumstances: one-way interaction and two-way interaction.

In the partial conditions, some of the communication is only in one direction, so that any deliberate adjustment in harmonic interval must be attributed to the singer who can hear their partner. In this case, we have a *one-way interaction*. In the open conditions, both singers in a pair are able to adjust to each other, creating a *two-way interaction*. Taking soprano isolated conditions as an example, the harmonic intervals involving soprano are one-way interactions, and the harmonic intervals between alto, tenor and bass are two-way interactions (Figure 3).

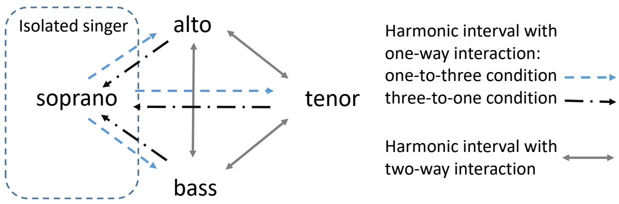

Figure 3: Interaction in the soprano isolated conditions

We compare the MAHIE for two-way interactions with those for one-way interactions in the three-to-one test conditions. MAHIE is significantly smaller for the two-way interactions than for one-way interactions (F(1,23659)=10.94, p<.001). This supports the third hypothesis, and indicates that acoustic feedback helps singers to interactively tune harmonic intervals.

However, no significant difference was found between MAHIE for different directions of intonation, that is the three-to-one condition versus the one-to-three condition (F(1,23524)=0.39, p=0.53). When one side of interactive intonation is without acoustic feedback, the direction of the feedback does not appear to influence the harmonic interval.

4.4 Note Stability

The note stability is measured by its converse, note variability (Eq. 7). The acoustic feedback of other singers not only has an influence on intonation accuracy (section 4.2) but also has an influence on note variability.

The note variability in the closed condition is significantly different from that in the partial and open conditions (F(1,23659)=41.23, p<.001), but no significant difference was found between the partial and open conditions (F(1,22514)=1.37, p=0.24). Note trajectories become less stable when singers can hear other singers in addition to their own voice, which is further evidence of interaction in intonation. This agrees with previous studies, which show that singers perform worse when singing with an unstable reference pitch [4, 16].

Moreover, the note variability is weakly positively correlated to the MAPE of individual notes (r=0.18, $p<.001$), but it is not obviously related to the singer (r=0.01, p=0.01) or training experience (r=0.08, $p<.001$).

The fourth hypothesis has been tested, and the results confirm that there is a relationship between the listening condition and note stability. This complements results from other research which assert that note stability of individual singers depends on emotional expression [5, 21]. Other possible relationships, such as a connection between musical training and note stability, were not supported by the experimental results.

5. DISCUSSION AND FUTURE WORK

This study tested four hypotheses using various metrics of singing accuracy and statistical tests. In each case, significant results were found. In three of the four cases, the results supported the hypotheses, however for the first hypothesis, the direction of the observed effect was the opposite of what was predicted.

Participants noted that the bass part (male singer) is the most difficult vocal part to recruit. It is possible that this leads to a lower average standard among bass singers. A comparison of pitch error by vocal type reveals that the bass vocal part has a larger MAPE than the other vocal parts. This may be the cause of the unexpected result for the bass isolated condition: i.e. because the bass voice had greater pitch error, other parts which tuned to the bass also increased their pitch error.

The factor of interaction, that is when singers can hear each other, increases the pitch error of the individual singers but decreases the harmonic interval error between the singers. Although these results may appear to be contradictory, this can occur when interval errors accumulate, and the sung pitches drift away from the initial tonal reference, as has been demonstrated by Howard [6].

Many factors of influence have been researched which are crucial for singing, such as age and gender (boys are more likely to sing out of tune than girls), and individual differences [27]. As it is not possible to cover all aspects in this paper, we leave the analysis of results from the questionnaire to future work, including the investigation of the relationships between intonation accuracy and active engagement with music, perceptual abilities, musical training and singing ability.

6. CONCLUSIONS

For analysis of the effect of interaction on intonation in unaccompanied SATB singing, we designed a novel experiment and tested the intonation accuracy of five groups of singers in a series of test and listening conditions. The results confirm that interaction exists between singers and influences their intonation, and that intonation accuracy depends on which other singers each individual singer can hear.

In particular, we observed that the three-to-one bass isolated test condition had a significantly lower MAPE compared with other three-to-one conditions, and compared with the open condition. In other words, singers were more accurate when they could not hear the bass. This surprising result might be due to the fact that the bass singers were less accurate on average than other singers in this experiment.

We observed increases in pitch error and melodic interval error when singers could hear each other. The closed condition had the smallest MAMIE, while the open condition had the largest. At the same time, acoustic feedback decreased the harmonic interval error, while the direction of the feedback did not influence the harmonic interval error.

Interaction also has the effect of reducing the note stability, or increasing its variability. Pitch within a note varies more when singers hear each other, as one might expect if the singers are adjusting their intonation to be in tune with each other.

In conclusion, this paper addresses a gap in singing intonation studies, by investigating the effects of interaction between singers. We found that interaction significantly influences the pitch accuracy, leading to increases in the pitch error, melodic interval error, and note stability but a decrease in the harmonic interval error. Although many aspects of the data remain to be explored, we hope the current results provide useful information and better understanding of interactive intonation.

7. ACKNOWLEDGEMENT

Many thanks to all of the participants who contributed to this project, including the QMUL *A Capella* Society, QMUL Music Society, London Philharmonic Choir, the Hi-Fan Vocal Group. We also thank Marcus Pearce, Daniel Stowell and Christophe Rhodes for their advice on experimental design. Jiajie Dai is supported by a China Scholarship Council and Queen Mary Joint PhD Scholarship.

8. DATA AVAILABILITY

Annotated data, experimental score and code to reproduce our results are available at: `https://code.soundsoftware.ac.uk/projects/analysis-of-interactive-intonation-in-unaccompanied-satb-ensembles/repository`.

9. REFERENCES

[1] Per-Gunnar Alldahl. *Choral Intonation*. Gehrmans, 2008.

[2] Jocelei C S Bohrer. Intonational Strategies in Ensemble Singing. *Doctoral thesis, City University London*, 2002.

[3] Brian J Brandler and Zehra F Peynircioglu. A Comparison of the Efficacy of Individual and Collaborative Music Learning in Ensemble Rehearsals. *Journal of Research in Music Education*, 63(3):281–297, 2015.

[4] Jiajie Dai and Simon Dixon. Analysis of Vocal Imitation of Pitch Trajectories. *17th ISMIR Conference*, 2016.

[5] Janina Fyk. *Melodic Intonation, Psychoacoustics and the Violin*. Organon, 1995.

[6] David M Howard. Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice*, 21(3):300–315, 2007.

[7] Joyce B Kennedy and Michael Kennedy. *The Concise Oxford Dictionary of Music*. Oxford University Press, 2004.

[8] Joan La Barbara. Voice is the Original Instrument. *Contemporary Music Review*, 21(1):35–48, 2002.

[9] Peggy A Long. Relationships Between Pitch Memory in Short Melodies and Selected Factors. *Journal of Research in Music Education*, 25(4):272–282, 1977.

[10] M. Mauch, C. Cannam, R. Bittner, G. Fazekas, J. Bello J. Salamon and, J. Dai, and S. Dixon. Tony: a Tool for Efficient Computer-aided Melody Note Transcription. In *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR)*, 2015.

[11] M. Mauch and S. Dixon. PYIN: a fundamental frequency estimator using probabilistic threshold distributions. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 659–663, 2014.

[12] Matthias Mauch, Klaus Frieler, and Simon Dixon. Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory. *The Journal of the Acoustical Society of America*, 136(1):401–411, 2014.

[13] Steven J Mithen. *The Singing Neanderthal: A Search for the Origins of Art, Religion, and Science*. Harvard University Press, Cambridge, MA, 2007. esp. Ch. 16, pp. 246–265.

[14] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642, 2014.

[15] Dirk Mürbe, Friedemann Pabst, Gert Hofmann, and Johan Sundberg. Significance of Auditory and Kinesthetic Feedback to Singers' Pitch Control. *Journal of Voice*, 16(1):44–51, 2002.

[16] Peter Q Pfordresher and Steven Brown. Poor-pitch Singing in The Absence of "Tone Deafness". *Music Perception: An Interdisciplinary Journal*, 25(2):95–115, 2007.

[17] Peter Q Pfordresher, Steven Brown, Kimberly M Meier, Michel Belyk, and Mario Liotti. Imprecise Singing is Widespread. *The Journal of the Acoustical Society of America*, 128(4):2182–2190, 2010.

[18] John Potter, editor. *The Cambridge Companion to Singing*. Cambridge University Press, 2000.

[19] E. Prout. *Harmony: Its Theory and Practice*. Cambridge University Press, 2011.

[20] R. Seaton, D. Pim, and D. Sharp. Pitch Drift in a Cappella Choral Singing-Work in Progress Report. *Institute of Acoustics Annual Spring Conference 2013*, 2013.

[21] Johan Sundberg, Filipa M B Lã, and Evangelos Himonides. Intonation and Expressivity: A Single Case Study of Classical Western Singing. *Journal of Voice*, 27(3):391.e1–e8, 2013.

[22] Annie H Takeuchi and Stewart H Hulse. Absolute pitch. *Psychological Bulletin*, 113(2):345–361, 1993.

[23] Hiroko Terasawa. Pitch drift in choral music, 2004. Music 221A final paper, Center for Computer Research in Music and Acoustics, Stanford University, CA.

[24] Sten Ternström and Johan Sundberg. Intonation Precision of Choir Singers. *The Journal of the Acoustical Society of America*, 84(1):59–69, 1988.

[25] Richard A Warren and Meagan E Curtis. The Actual vs. Predicted Effects of Intonation Accuracy on Vocal Performance Quality. *Music Perception: An Interdisciplinary Journal*, 33(2):135–146, 2015.

[26] Graham F Welch. Singing as communication. *Musical Communication*, pages 239–259, 2005.

[27] Graham F Welch, Desmond C Sergeant, and Peta J White. Age, Sex, and Vocal Task as Factors In Singing "in Tune" During The First Years of Schooling. *Bulletin of the Council for Research in Music Education*, pages 153–160, 1997.

[28] Vivienne M Young and Andrew M Colman. Some Psychological Processes in String Quartets. *Psychology of Music*, 7(1):12–18, 1979.